

Audio/Video Transport WG  
Internet-Draft  
Intended status: Standards Track  
Expires: May 23, 2011

T. Kristensen  
P. Luthi  
TANDBERG  
November 19, 2010

**RTP Payload Format for H.264 Reduced-Complexity Decoding Operation  
(RCDO) Video  
draft-ietf-avt-rtp-h264-rcdo-08**

**Abstract**

This document describes an RTP payload format for the Reduced-Complexity Decoding Operation (RCDO) for H.264 Baseline profile bitstreams, as specified in ITU-T Recommendation H.241. RCDO reduces the decoding cost and resource consumption of the video processing. The RCDO RTP payload format is based on the H.264 RTP payload format.

**Status of this Memo**

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on May 23, 2011.

**Copyright Notice**

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as

described in the Simplified BSD License.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

## Table of Contents

<a href="#">1.</a>	Introduction . . . . .	<a href="#">3</a>
<a href="#">2.</a>	Conventions, Definitions and Acronyms . . . . .	<a href="#">3</a>
<a href="#">3.</a>	Media Format Background . . . . .	<a href="#">3</a>
<a href="#">4.</a>	Payload Format . . . . .	<a href="#">4</a>
<a href="#">5.</a>	Congestion Control Considerations . . . . .	<a href="#">4</a>
<a href="#">6.</a>	Payload Format Parameters . . . . .	<a href="#">4</a>
<a href="#">6.1.</a>	Media Type Definition . . . . .	<a href="#">4</a>
<a href="#">7.</a>	Mapping to SDP . . . . .	<a href="#">20</a>
<a href="#">7.1.</a>	Offer/Answer Considerations . . . . .	<a href="#">20</a>
<a href="#">7.2.</a>	Declarative SDP Considerations . . . . .	<a href="#">20</a>
<a href="#">8.</a>	IANA Considerations . . . . .	<a href="#">21</a>
<a href="#">9.</a>	Security Considerations . . . . .	<a href="#">21</a>
<a href="#">10.</a>	Acknowledgements . . . . .	<a href="#">21</a>
<a href="#">11.</a>	References . . . . .	<a href="#">21</a>
<a href="#">11.1.</a>	Normative References . . . . .	<a href="#">21</a>
<a href="#">11.2.</a>	Informative references . . . . .	<a href="#">22</a>
	Authors' Addresses . . . . .	<a href="#">22</a>



## **1. Introduction**

ITU-T Recommendation H.241 [3] specifies a reduced-complexity decoding operation (RCDO) for use with H.264 [2] Baseline profile bitstreams. It also specifies a bitstream constraint associated with RCDO and a mechanism for signalling RCDO within the bitstream. The RCDO signalling indicates that the bitstream conforms to the bitstream constraint and that the decoder shall apply the RCDO decoding process to the bitstream.

RCDO for H.264 offers a solution to support higher resolutions at the same high framerates used in current implementations. This is achieved by reducing the processing requirements and thus the decoding cost/resource consumption of the video processing.

This document defines media type parameters and allows use in systems based on the Session Description Protocol (SDP) [8] for signalling.

## **2. Conventions, Definitions and Acronyms**

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [4].

RFC-editor note: RFC XXXX is to be replaced by the RFC number this specification receives when published.

RFC-editor note: RFC YYYY is to be replaced by the RFC number the normative reference RFC3984bis [1] receives when published.

## **3. Media Format Background**

The Reduced-Complexity Decoding Operation (RCDO) for H.264 Baseline profile bitstreams is specified in Annex B of H.241 [3]. RCDO is specified as a separate H.264 mode, and is distinct from any profile defined in H.264. An RCDO bitstream obey to all the constraints of the Baseline profile.

The media format is based on the H.264 RTP payload format as specified in RFC YYYY [1]. Therefore, RFC YYYY constitutes the basis for this document and is referred to several times.

In order to signal H.264 additional modes, Table 9f of H.241 [3] specifies an AdditionalModesSupported parameter. Currently, the only additional mode defined is RCDO.



Informative note: Other additional modes may be defined in the future. H.264 additional modes may or may not be distinct from the Profiles in H.264.

A separate media subtype, named H264-RCDO, is defined to ensure backward compatibility with deployed implementations of H.264.

#### **4. Payload Format**

The payload format defined in [Section 5](#) of RFC YYYY [1] SHALL be used. This includes the RTP header usage and the payload format in RFC YYYY. Examples of typical RTP packets can be found in RFC YYYY.

#### **5. Congestion Control Considerations**

Congestion control for RTP SHALL be used in accordance with [RFC 3550](#) [6], and with any applicable RTP profile; e.g., [RFC 3551](#) [7]. If best-effort service is being used, users of this payload format SHALL monitor packet loss to ensure that the packet loss rate is within acceptable parameters.

#### **6. Payload Format Parameters**

This RTP payload format is identified using the H264-RCDO media subtype which is registered in accordance with [RFC 4855](#) [10] and using the template of [RFC 4288](#) [13].

##### **6.1. Media Type Definition**

RFC-editor note: We need to sync with RFC YYYY [1], when it is certain no more changes will go into that draft, and copy across any changes to the identical part of the media subtype definition.

Informative note: The media subtype definition for H264-RCDO is based on the definition of the H264 media subtype as specified in [Section 8.1](#) of RFC YYYY [1]. Except for the profile-level-id parameter, where new semantics are specified below, the optional parameters are copied from RFC YYYY [1] in order to provide a complete, self-contained media subtype registration to IANA. The references are updated to match the numbering used in this document.

The media subtype for RCDO for H.264 is allocated from the IETF tree.

Type name: video



Subtype name: H264-RCD0

Required parameters:

rate: Indicates the RTP timestamp clock rate. The rate value MUST be 90000.

Optional parameters:

profile-level-id: A base16 [RFC 4648](#) [9] (hexadecimal) representation of the following three bytes in the sequence parameter set NAL unit specified in H.264 [2]: 1) profile\_idc, 2) a byte herein referred to as profile-iop, composed of the values of constraint\_set0\_flag, constraint\_set1\_flag, constraint\_set2\_flag, constraint\_set3\_flag, and reserved\_zero\_4bits in bit-significance order, starting from the most significant bit, and 3) level\_idc. Note that reserved\_zero\_4bits is required to be equal to 0 in H.264 [2], but other values for it may be specified in the future by ITU-T or ISO/IEC.

The profile-level-id parameter indicates the default sub-profile, i.e. the subset of coding tools that may have been used to generate the stream or that the receiver supports, and the default level of the stream or the receiver supports.

RCD0 is distinct from any profile, this implies that the profile value 0 (no profile) and the profile\_idc byte of the profile-level-id parameter are equal to 0. An RCD0 bitstream MUST obey to all the constraints of the Baseline profile. Therefore, only constraint\_set0\_flag is equal to 1 in the profile-iop part of the profile-level-id parameter, the remaining bits are set to 0.

If the profile-level-id parameter is used to indicate properties of a NAL unit stream, it indicates that, to decode the stream, the lowest level the decoder has to support is the default level. If the profile-level-id parameter is used for capability exchange or session setup procedure, and if max-recv-level is not present, the default level from profile-level-id indicates the highest level the codec wishes to support. If max-recv-level is present it indicates the highest level the codec supports for receiving. For either receiving or sending, all levels that are lower than the highest level supported MUST also be supported.

For example, if a codec supports level 1.3, the profile-level-id becomes 00800d, in which 00 indicates the "no profile" value, 80 indicates the constraints of the Baseline profile and 0d indicates level 1.3. When level 2.1 is supported, the profile-level-id becomes 008015.





If no profile-level-id is present, level 1 MUST be implied, i.e. equivalent to profile-level-id 00800a.

Informative note: The definitions of the remaining optional parameters below are copied verbatim from [Section 8.1](#) of RFC YYYY [1]. Only the references are updated to match the numbering used in this document.

max-recv-level: This parameter MAY be used to indicate the highest level a receiver supports when the highest level is higher than the default level (the level indicated by profile-level-id). The value of max-recv-level is a base16 (hexadecimal) representation of the two bytes after the syntax element profile\_idc in the sequence parameter set NAL unit specified in H.264 [2]: profile\_iop (as defined above) and level\_idc. If (the level\_idc byte of max-recv-level is equal to 11 and bit 4 of the profile\_iop byte of max-recv-level is equal to 1) or (the level\_idc byte of max-recv-level is equal to 9 and bit 4 of the profile\_iop byte of max-recv-level is equal to 0), the highest level the receiver supports is level 1b. Otherwise, the highest level the receiver supports is equal to the level\_idc byte of max-recv-level divided by 10.

max-recv-level MUST NOT be present if the highest level the receiver supports is not higher than the default level.

max-mbps, max-smbps, max-fs, max-cpb, max-dpb, and max-br: These parameters MAY be used to signal the capabilities of a receiver implementation. These parameters MUST NOT be used for any other purpose. The highest level conveyed in the value of the profile-level-id parameter or the max-recv-level parameter MUST be such that the receiver is fully capable of supporting. max-mbps, max-smbps, max-fs, max-cpb, max-dpb, and max-br MAY be used to indicate capabilities of the receiver that extend the required capabilities of the signaled highest level, as specified below.

When more than one parameter from the set (max-mbps, max-smbps, max-fs, max-cpb, max-dpb, max-br) is present, the receiver MUST support all signaled capabilities simultaneously. For example, if both max-mbps and max-br are present, the signaled highest level with the extension of both the frame rate and bit rate is supported. That is, the receiver is able to decode NAL unit streams in which the macroblock processing rate is up to max-mbps (inclusive), the bit rate is up to max-br (inclusive), the coded picture buffer size is derived as specified in the semantics of the max-br parameter below, and other properties comply with the highest level specified in the value of the profile-level-id parameter or the max-recv-level parameter.



If a receiver can support all the properties of level A, the highest level specified in the value of the profile-level-id parameter or the max-recv-level parameter MUST be level A (i.e. MUST NOT be lower than level A). In other words, a receiver MUST NOT signal values of max-mbps, max-fs, max-cpb, max-dpb, and max-br that taken together meet the requirements of a higher level compared to the highest level specified in the value of the profile-level-id parameter or the max-recv-level parameter.

Informative note: When the OPTIONAL media type parameters are used to signal the properties of a NAL unit stream, max-mbps, max-smbps, max-fs, max-cpb, max-dpb, and max-br are not present, and the value of profile-level-id must always be such that the NAL unit stream complies fully with the specified profile and level.

max-mbps: The value of max-mbps is an integer indicating the maximum macroblock processing rate in units of macroblocks per second. The max-mbps parameter signals that the receiver is capable of decoding video at a higher rate than is required by the signaled highest level conveyed in the value of the profile-level-id parameter or the max-recv-level parameter. When max-mbps is signaled, the receiver MUST be able to decode NAL unit streams that conform to the signaled highest level, with the exception that the MaxMBPS value in Table A-1 of H.264 [2] for the signaled highest level is replaced with the value of max-mbps. The value of max-mbps MUST be greater than or equal to the value of MaxMBPS given in Table A-1 of H.264 [2] for the highest level. Senders MAY use this knowledge to send pictures of a given size at a higher picture rate than is indicated in the signaled highest level.

max-smbps: The value of max-smbps is an integer indicating the maximum static macroblock processing rate in units of static macroblocks per second, under the hypothetical assumption that all macroblocks are static macroblocks. When max-smbps is signalled the MaxMBPS value in Table A-1 of H.264 [2] should be replaced with the result of the following computation:

- o If the parameter max-mbps is signalled, set a variable MaxMacroblocksPerSecond to the value of max-mbps. Otherwise, set MaxMacroblocksPerSecond equal to the value of MaxMBPS in Table A-1 H.264 [2] for the highest level.



- o Set a variable `P_non-static` to the proportion of non- static macroblocks in picture `n`.
- o Set a variable `P_static` to the proportion of static macroblocks in picture `n`.
- o The value of `MaxMBPS` in Table A-1 of H.264 [2] should be considered by the encoder to be equal to:

$$\text{MaxMacroblocksPerSecond} * \text{max-smbps} / (\text{P\_non-static} * \text{max-smbps} + \text{P\_static} * \text{MaxMacroblocksPerSecond})$$

The encoder should recompute this value for each picture. The value of `max-smbps` MUST be greater than the value of `MaxMBPS` given in Table A-1 of H.264 [2] for the highest level. Senders MAY use this knowledge to send pictures of a given size at a higher picture rate than is indicated in the signaled highest level.

`max-fs`: The value of `max-fs` is an integer indicating the maximum frame size in units of macroblocks. The `max-fs` parameter signals that the receiver is capable of decoding larger picture sizes than are required by the signaled highest level conveyed in the value of the `profile-level-id` parameter or the `max-recv-level` parameter. When `max-fs` is signaled, the receiver MUST be able to decode NAL unit streams that conform to the signaled highest level, with the exception that the `MaxFS` value in Table A-1 of H.264 [2] for the signaled highest level is replaced with the value of `max-fs`. The value of `max-fs` MUST be greater than or equal to the value of `MaxFS` given in Table A-1 of H.264 [2] for the highest level. Senders MAY use this knowledge to send larger pictures at a proportionally lower frame rate than is indicated in the signaled highest level.

`max-cpb`: The value of `max-cpb` is an integer indicating the maximum coded picture buffer size in units of 1000 bits for the VCL HRD parameters (see A.3.1 item i of H.264 [2]) and in units of 1200 bits for the NAL HRD parameters (see A.3.1 item j of H.264 [2]). The `max-cpb` parameter signals that the receiver has more memory than the minimum amount of coded picture buffer memory required by the signaled highest level conveyed in the value of the `profile-level-id` parameter or the `max-recv-level` parameter. When `max-cpb` is signaled, the receiver MUST be able to decode NAL unit streams that conform to the signaled highest level, with the exception that the `MaxCPB` value in Table A-1 of H.264 [2] for the signaled highest level is replaced with the value of `max-cpb`. The value of `max-cpb` MUST be greater than or equal to the value of `MaxCPB` given in Table A-1 of H.264 [2] for the highest level. Senders MAY use



this knowledge to construct coded video streams with greater variation of bit rate than can be achieved with the MaxCPB value in Table A-1 of H.264 [2].

Informative note: The coded picture buffer is used in the hypothetical reference decoder (Annex C) of H.264. The use of the hypothetical reference decoder is recommended in H.264 encoders to verify that the produced bitstream conforms to the standard and to control the output bitrate. Thus, the coded picture buffer is conceptually independent of any other potential buffers in the receiver, including de-interleaving and de-jitter buffers. The coded picture buffer need not be implemented in decoders as specified in Annex C of H.264, but rather standard-compliant decoders can have any buffering arrangements provided that they can decode standard-compliant bitstreams. Thus, in practice, the input buffer for video decoder can be integrated with de-interleaving and de-jitter buffers of the receiver.

max-dpb: The value of max-dpb is an integer indicating the maximum decoded picture buffer size in units of 1024 bytes. The max-dpb parameter signals that the receiver has more memory than the minimum amount of decoded picture buffer memory required by the signaled highest level conveyed in the value of the profile-level-id parameter or the max-recv-level parameter. When max-dpb is signaled, the receiver MUST be able to decode NAL unit streams that conform to the signaled highest level, with the exception that the MaxDPB value in Table A-1 of H.264 [2] for the signaled highest level is replaced with the value of max-dpb. Consequently, a receiver that signals max-dpb MUST be capable of storing the following number of decoded frames, complementary field pairs, and non-paired fields in its decoded picture buffer:

$$\text{Min}(1024 * \text{max-dpb} / (\text{PicWidthInMbs} * \text{FrameHeightInMbs} * 256 * \text{ChromaFormatFactor}), 16)$$

PicWidthInMbs, FrameHeightInMbs, and ChromaFormatFactor are defined in H.264 [2].

The value of max-dpb MUST be greater than or equal to the value of MaxDPB given in Table A-1 of H.264 [2] for the highest level. Senders MAY use this knowledge to construct coded video streams with improved compression.





Informative note: This parameter was added primarily to complement a similar codepoint in the ITU-T Recommendation H.245, so as to facilitate signaling gateway designs. The decoded picture buffer stores reconstructed samples. There is no relationship between the size of the decoded picture buffer and the buffers used in RTP, especially de-interleaving and de-jitter buffers.

max-br: The value of max-br is an integer indicating the maximum video bit rate in units of 1000 bits per second for the VCL HRD parameters (see A.3.1 item i of H.264 [2]) and in units of 1200 bits per second for the NAL HRD parameters (see A.3.1 item j of H.264 [2]).

The max-br parameter signals that the video decoder of the receiver is capable of decoding video at a higher bit rate than is required by the signaled highest level conveyed in the value of the profile-level-id parameter or the max-recv-level parameter.

When max-br is signaled, the video codec of the receiver MUST be able to decode NAL unit streams that conform to the signaled highest level, with the following exceptions in the limits specified by the highest level:

- o The value of max-br replaces the MaxBR value in Table A-1 of H.264 [2] for the highest level.
- o When the max-cpb parameter is not present, the result of the following formula replaces the value of MaxCPB in Table A-1 of H.264 [2]:  $(\text{MaxCPB of the signaled level}) * \text{max-br} / (\text{MaxBR of the signaled highest level})$ .

For example, if a receiver signals capability for Level 1.2 with max-br equal to 1550, this indicates a maximum video bitrate of 1550 kbits/sec for VCL HRD parameters, a maximum video bitrate of 1860 kbits/sec for NAL HRD parameters, and a CPB size of 4036458 bits  $(1550000 / 384000 * 1000 * 1000)$ .

The value of max-br MUST be greater than or equal to the value MaxBR given in Table A-1 of H.264 [2] for the signaled highest level.

Senders MAY use this knowledge to send higher bitrate video as allowed in the level definition of Annex A of H.264, to achieve improved video quality.



Informative note: This parameter was added primarily to complement a similar codepoint in the ITU-T Recommendation H.245, so as to facilitate signaling gateway designs. No assumption can be made from the value of this parameter that the network is capable of handling such bit rates at any given time. In particular, no conclusion can be drawn that the signaled bit rate is possible under congestion control constraints.

**redundant-pic-cap:** This parameter signals the capabilities of a receiver implementation. When equal to 0, the parameter indicates that the receiver makes no attempt to use redundant coded pictures to correct incorrectly decoded primary coded pictures. When equal to 0, the receiver is not capable of using redundant slices; therefore, a sender **SHOULD** avoid sending redundant slices to save bandwidth. When equal to 1, the receiver is capable of decoding any such redundant slice that covers a corrupted area in a primary decoded picture (at least partly), and therefore a sender **MAY** send redundant slices. When the parameter is not present, then a value of 0 **MUST** be used for redundant-pic-cap. When present, the value of redundant-pic-cap **MUST** be either 0 or 1.

When the profile-level-id parameter is present in the same signaling as the redundant-pic-cap parameter, and the profile indicated in profile-level-id is such that it disallows the use of redundant coded pictures (e.g., Main Profile), the value of redundant-pic-cap **MUST** be equal to 0. When a receiver indicates redundant-pic-cap equal to 0, the received stream **SHOULD NOT** contain redundant coded pictures.

Informative note: Even if redundant-pic-cap is equal to 0, the decoder is able to ignore redundant codec pictures provided that the decoder supports such a profile (Baseline, Extended) in which redundant coded pictures are allowed.

Informative note: Even if redundant-pic-cap is equal to 1, the receiver may also choose other error concealment strategies to replace or complement decoding of redundant slices.

**sprop-parameter-sets:** This parameter **MAY** be used to convey any sequence and picture parameter set NAL units (herein referred to as the initial parameter set NAL units) that can be placed in the NAL unit stream to precede any other NAL units in decoding order. The parameter **MUST NOT** be used to indicate codec capability in any capability exchange procedure. The value of the parameter is a comma (',') separated list of base64 [RFC 4648](#) [9] representations



of parameter set NAL units as specified in sections [7.3.2.1](#) and 7.3.2.2 of H.264 [2]. Note that the number of bytes in a parameter set NAL unit is typically less than 10, but a picture parameter set NAL unit can contain several hundreds of bytes.

Informative note: When several payload types are offered in the SDP Offer/Answer model, each with its own sprop- parameter-sets parameter, then the receiver cannot assume that those parameter sets do not use conflicting storage locations (i.e., identical values of parameter set identifiers). Therefore, a receiver should buffer all sprop-parameter-sets and make them available to the decoder instance that decodes a certain payload type.

The "sprop-parameter-sets" parameter MUST only contain parameter sets that are conforming to the profile-level-id, i.e., the subset of coding tools indicated by any of the parameter sets MUST be equal to the default sub-profile, and the level indicated by any of the parameter sets MUST be equal to the default level.

sprop-level-parameter-sets: This parameter MAY be used to convey any sequence and picture parameter set NAL units (herein referred to as the initial parameter set NAL units) that can be placed in the NAL unit stream to precede any other NAL units in decoding order and that are associated with one or more levels different than the default level. The parameter MUST NOT be used to indicate codec capability in any capability exchange procedure.

The sprop-level-parameter-sets parameter contains parameter sets for one or more levels which are different than the default level. All parameter sets associated with one level are clustered and prefixed with a three-byte field which has the same syntax as profile-level-id. This enables the receiver to install the parameter sets for one level and discard the rest. The three-byte field is named PLId, and all parameter sets associated with one level are named PSL, which has the same syntax as sprop-parameter-sets. Parameter sets for each level are represented in the form of PLId:PSL, i.e., PLId followed by a colon (':') and the base64 [RFC 4648](#) [9] representation of the initial parameter set NAL units for the level. Each pair of PLId:PSL is also separated by a colon. Note that a PSL can contain multiple parameter sets for that level, separated with commas (',').

The subset of coding tools indicated by each PLId field MUST be equal to the default sub-profile, and the level indicated by each PLId field MUST be different than the default level. All sequence parameter sets contained in each PSL MUST have the three bytes from profile\_idc to level\_idc, inclusive, equal to the preceding



PLId.

Informative note: This parameter allows for efficient level downgrade or upgrade in SDP Offer/Answer and out- of-band transport of parameter sets, simultaneously.

use-level-src-parameter-sets: This parameter MAY be used to indicate a receiver capability. The value MAY be equal to either 0 or 1. When the parameter is not present, the value MUST be inferred to be equal to 0. The value 0 indicates that the receiver does not understand the sprop-level-parameter-sets parameter, and does not understand the "fntp" source attribute as specified in [section 6.3 of RFC 5576 \[14\]](#), and will ignore sprop-level-parameter- sets when present, and will ignore sprop-parameter-sets when conveyed using the "fntp" source attribute. The value 1 indicates that the receiver understands the sprop-level- parameter-sets parameter, and understands the "fntp" source attribute as specified in [section 6.3 of RFC 5576 \[14\]](#), and is capable of using parameter sets contained in the sprop-level- parameter-sets or contained in the sprop-parameter-sets that is conveyed using the "fntp" source attribute.

Informative note: An [RFC 3984](#) receiver does not understand sprop-level-parameter-sets, use-level-src- parameter-sets, or the "fntp" source attribute as specified in section 6.3 of [RFC 5576 \[14\]](#). Therefore, during SDP Offer/Answer, an [RFC 3984](#) receiver as the answerer will simply ignore sprop-level- parameter-sets, when present in an offer, and sprop-parameter-sets conveyed using the "fntp" source attribute as specified in [section 6.3 of RFC 5576 \[14\]](#). Assume that the offered payload type was accepted at a level lower than the default level. If the offered payload type included sprop-level-parameter-sets or included sprop-parameter-sets conveyed using the "fntp" source attribute, and the offerer sees that the answerer has not included use-level-src-parameter-sets equal to 1 in the answer, the offerer knows that in-band transport of parameter sets is needed.

in-band-parameter-sets: This parameter MAY be used to indicate a receiver capability. The value MAY be equal to either 0 or 1. The value 1 indicates that the receiver discards out-of-band parameter sets in sprop-parameter-sets and sprop-level-parameter-sets, therefore the sender MUST transmit all parameter sets in-band. The value 0 indicates that the receiver utilizes out-of-band parameter sets included in sprop-parameter-sets and/or sprop-level-parameter-sets. However, in this case, the sender MAY still choose to send parameter sets in-band. When in-band- parameter-sets is equal to 1, use-level-src-parameter-sets MUST NOT be





present or MUST be equal to 0. When the parameter is not present, this receiver capability is not specified, and therefore the sender MAY send out-of-band parameter sets only, or it MAY send in-band-parameter-sets only, or it MAY send both.

**level-asymmetry-allowed:** This parameter MAY be used in SDP Offer/Answer to indicate whether level asymmetry, i.e., sending media encoded at a different level in the offerer-to-answerer direction than the level in the answerer-to-offerer direction, is allowed. The value MAY be equal to either 0 or 1. When the parameter is not present, the value MUST be inferred to be equal to 0. The value 1 in both the offer and the answer indicates that level asymmetry is allowed. The value of 0 in either the offer or the answer indicates the level asymmetry is not allowed.

If "level-asymmetry-allowed" is equal to 0 (or not present) in either the offer or the answer, level asymmetry is not allowed. In this case, the level to use in the direction from the offerer to the answerer MUST be the same as the level to use in the opposite direction.

**packetization-mode:** This parameter signals the properties of an RTP payload type or the capabilities of a receiver implementation. Only a single configuration point can be indicated; thus, when capabilities to support more than one packetization-mode are declared, multiple configuration points (RTP payload types) must be used.

When the value of packetization-mode is equal to 0 or packetization-mode is not present, the single NAL mode MUST be used. This mode is in use in standards using ITU-T Recommendation H.241 [3] (see [section 12.1](#)). When the value of packetization-mode is equal to 1, the non-interleaved mode MUST be used. When the value of packetization-mode is equal to 2, the interleaved mode MUST be used. The value of packetization-mode MUST be an integer in the range of 0 to 2, inclusive.

**sprop-interleaving-depth:** This parameter MUST NOT be present when packetization-mode is not present or the value of packetization-mode is equal to 0 or 1. This parameter MUST be present when the value of packetization-mode is equal to 2.

This parameter signals the properties of an RTP packet stream. It specifies the maximum number of VCL NAL units that precede any VCL NAL unit in the RTP packet stream in transmission order and follow the VCL NAL unit in decoding order. Consequently, it is guaranteed that receivers can reconstruct NAL unit decoding order when the buffer size for NAL unit decoding order recovery is at



least the value of `sprop-interleaving-depth + 1` in terms of VCL NAL units.

The value of `sprop-interleaving-depth` MUST be an integer in the range of 0 to 32767, inclusive.

`sprop-deint-buf-req`: This parameter MUST NOT be present when `packetization-mode` is not present or the value of `packetization-mode` is equal to 0 or 1. It MUST be present when the value of `packetization-mode` is equal to 2.

`sprop-deint-buf-req` signals the required size of the de-interleaving buffer for the RTP packet stream. The value of the parameter MUST be greater than or equal to the maximum buffer occupancy (in units of bytes) required in such a de-interleaving buffer that is specified in [section 7.2](#) of RFC YYYY [1]. It is guaranteed that receivers can perform the de-interleaving of interleaved NAL units into NAL unit decoding order, when the de-interleaving buffer size is at least the value of `sprop-deint-buf-req` in terms of bytes.

The value of `sprop-deint-buf-req` MUST be an integer in the range of 0 to 4294967295, inclusive.

Informative note: `sprop-deint-buf-req` indicates the required size of the de-interleaving buffer only. When network jitter can occur, an appropriately sized jitter buffer has to be provisioned for as well.

`deint-buf-cap`: This parameter signals the capabilities of a receiver implementation and indicates the amount of de-interleaving buffer space in units of bytes that the receiver has available for reconstructing the NAL unit decoding order. A receiver is able to handle any stream for which the value of the `sprop-deint-buf-req` parameter is smaller than or equal to this parameter.

If the parameter is not present, then a value of 0 MUST be used for `deint-buf-cap`. The value of `deint-buf-cap` MUST be an integer in the range of 0 to 4294967295, inclusive.

Informative note: `deint-buf-cap` indicates the maximum possible size of the de-interleaving buffer of the receiver only. When network jitter can occur, an appropriately sized jitter buffer has to be provisioned for as well.



sprop-init-buf-time: This parameter MAY be used to signal the properties of an RTP packet stream. The parameter MUST NOT be present, if the value of packetization-mode is equal to 0 or 1.

The parameter signals the initial buffering time that a receiver MUST wait before starting decoding to recover the NAL unit decoding order from the transmission order. The parameter is the maximum value of (decoding time of the NAL unit - transmission time of a NAL unit), assuming reliable and instantaneous transmission, the same timeline for transmission and decoding, and that decoding starts when the first packet arrives.

An example of specifying the value of sprop-init-buf-time follows. A NAL unit stream is sent in the following interleaved order, in which the value corresponds to the decoding time and the transmission order is from left to right:

0 2 1 3 5 4 6 8 7 ...

Assuming a steady transmission rate of NAL units, the transmission times are:

0 1 2 3 4 5 6 7 8 ...

Subtracting the decoding time from the transmission time column-wise results in the following series:

0 -1 1 0 -1 1 0 -1 1 ...

Thus, in terms of intervals of NAL unit transmission times, the value of sprop-init-buf-time in this example is 1. The parameter is coded as a non-negative base10 integer representation in clock ticks of a 90-kHz clock. If the parameter is not present, then no initial buffering time value is defined. Otherwise the value of sprop-init-buf-time MUST be an integer in the range of 0 to 4294967295, inclusive.

In addition to the signaled sprop-init-buf-time, receivers SHOULD take into account the transmission delay jitter buffering, including buffering for the delay jitter caused by mixers, translators, gateways, proxies, traffic-shapers, and other network elements.



sprop-max-don-diff: This parameter MAY be used to signal the properties of an RTP packet stream. It MUST NOT be used to signal transmitter or receiver or codec capabilities. The parameter MUST NOT be present if the value of packetization-mode is equal to 0 or 1. sprop-max-don-diff is an integer in the range of 0 to 32767, inclusive. If sprop-max-don-diff is not present, the value of the parameter is unspecified. sprop-max-don-diff is calculated as follows:

$$\text{sprop-max-don-diff} = \max\{\text{AbsDON}(i) - \text{AbsDON}(j)\}, \text{ for any } i \text{ and any } j > i,$$

where  $i$  and  $j$  indicate the index of the NAL unit in the transmission order and AbsDON denotes a decoding order number of the NAL unit that does not wrap around to 0 after 65535. In other words, AbsDON is calculated as follows: Let  $m$  and  $n$  be consecutive NAL units in transmission order. For the very first NAL unit in transmission order (whose index is 0),  $\text{AbsDON}(0) = \text{DON}(0)$ . For other NAL units, AbsDON is calculated as follows:

If  $\text{DON}(m) == \text{DON}(n)$ ,  $\text{AbsDON}(n) = \text{AbsDON}(m)$

If  $(\text{DON}(m) < \text{DON}(n) \text{ and } \text{DON}(n) - \text{DON}(m) < 32768)$ ,  
 $\text{AbsDON}(n) = \text{AbsDON}(m) + \text{DON}(n) - \text{DON}(m)$

If  $(\text{DON}(m) > \text{DON}(n) \text{ and } \text{DON}(m) - \text{DON}(n) \geq 32768)$ ,  
 $\text{AbsDON}(n) = \text{AbsDON}(m) + 65536 - \text{DON}(m) + \text{DON}(n)$

If  $(\text{DON}(m) < \text{DON}(n) \text{ and } \text{DON}(n) - \text{DON}(m) \geq 32768)$ ,  
 $\text{AbsDON}(n) = \text{AbsDON}(m) - (\text{DON}(m) + 65536 - \text{DON}(n))$

If  $(\text{DON}(m) > \text{DON}(n) \text{ and } \text{DON}(m) - \text{DON}(n) < 32768)$ ,  
 $\text{AbsDON}(n) = \text{AbsDON}(m) - (\text{DON}(m) - \text{DON}(n))$

where  $\text{DON}(i)$  is the decoding order number of the NAL unit having index  $i$  in the transmission order. The decoding order number is specified in [section 5.5](#) of RFC YYYY [1].

Informative note: Receivers may use sprop-max-don-diff to trigger which NAL units in the receiver buffer can be passed to the decoder.

max-rcmd-nalu-size: This parameter MAY be used to signal the capabilities of a receiver. The parameter MUST NOT be used for any other purposes. The value of the parameter indicates the largest NALU size in bytes that the receiver can handle efficiently. The parameter value is a recommendation, not a strict upper boundary. The sender MAY create larger NALUs but





must be aware that the handling of these may come at a higher cost than NALUs conforming to the limitation.

The value of max-rcmd-nalu-size MUST be an integer in the range of 0 to 4294967295, inclusive. If this parameter is not specified, no known limitation to the NALU size exists. Senders still have to consider the MTU size available between the sender and the receiver and SHOULD run MTU discovery for this purpose.

This parameter is motivated by, for example, an IP to H.223 video telephony gateway, where NALUs smaller than the H.223 transport data unit will be more efficient. A gateway may terminate IP; thus, MTU discovery will normally not work beyond the gateway.

Informative note: Setting this parameter to a lower than necessary value may have a negative impact.

sar-understood: This parameter MAY be used to indicate a receiver capability and not anything else. The parameter indicates the maximum value of aspect\_ratio\_idc (specified in H.264 [2]) smaller than 255 that the receiver understands. Table E-1 of H.264 [2] specifies aspect\_ratio\_idc equal to 0 as "unspecified", 1 to 16, inclusive, as specific Sample Aspect Ratios (SARs), 17 to 254, inclusive, as "reserved", and 255 as the Extended SAR, for which SAR width and SAR height are explicitly signaled. Therefore, a receiver with a decoder according to H.264 [2] understands aspect\_ratio\_idc in the range of 1 to 16, inclusive and aspect\_ratio\_idc equal to 255, in the sense that the receiver knows what exactly the SAR is. For such a receiver, the value of sar-understood is 16. If in the future Table E-1 of H.264 [2] is extended, e.g., such that the SAR for aspect\_ratio\_idc equal to 17 is specified, then for a receiver with a decoder that understands the extension, the value of sar-understood is 17. For a receiver with a decoder according to the 2003 version of H.264 [2], the value of sar-understood is 13, as the minimum reserved aspect\_ratio\_idc therein is 14.

When sar-understood is not present, the value MUST be inferred to be equal to 13.

sar-supported: This parameter MAY be used to indicate a receiver capability and not anything else. The value of this parameter is an integer in the range of 1 to sar-understood, inclusive, equal to 255. The value of sar-supported equal to N smaller than 255 indicates that the receiver supports all the SARs corresponding to H.264 aspect\_ratio\_idc values (see Table E-1 of H.264 [2]) in the range from 1 to N, inclusive, without geometric distortion. The value of sar-supported equal to 255 indicates that the receiver



supports all sample aspect ratios which are expressible using two 16-bit integer values as the numerator and denominator, i.e., those that are expressible using the H.264 `aspect_ratio_idc` value of 255 (Extended\_SAR, see Table E-1 of H.264 [2]), without geometric distortion.

H.264 compliant encoders SHOULD NOT send an `aspect_ratio_idc` equal to 0, or an `aspect_ratio_idc` larger than `sar-understood` and smaller than 255. H.264 compliant encoders SHOULD send an `aspect_ratio_idc` that the receiver is able to display without geometrical distortion. However, H.264 compliant encoders MAY choose to send pictures using any SAR.

Note that the actual sample aspect ratio or extended sample aspect ratio, when present, of the stream is conveyed in the Video Usability Information (VUI) part of the sequence parameter set.

Encoding considerations: This type is only defined for transfer via RTP ([RFC 3550](#)) and is framed and binary, see [section 4.8 in RFC4288](#).

Security considerations: See [section 9](#) of RFC XXXX.

Interoperability considerations: None

Published specification: RFC XXXX and its reference section.

Applications that use this media type: Video streaming and conferencing applications.

Additional information: None

Magic number(s):

File extension(s):

Macintosh file type code(s):

Person & email address to contact for further information:

Tom Kristensen <tom.kristensen@tandberg.com>, <tomkri@ifi.uio.no>

Intended usage: COMMON

Restrictions on usage: This type depends on RTP framing, and hence is only defined for transfer via RTP, ref [RFC3550](#). Transport within other framing protocols is not defined at this time.



Author: Tom Kristensen

Change controller: IETF Audio/Video Transport working group  
delegated from the IESG.

## 7. Mapping to SDP

The mapping of the above defined payload format media subtype and its parameters SHALL be done according to [Section 3 of RFC 4855](#) [10].

An example of the fmpdp attribute in the media representation of a level 2.2 bitstream is as follows:

```
a=fmpdp:97 profile-level-id=008016
```

### 7.1. Offer/Answer Considerations

When H264-RCDO is offered over RTP using SDP in an Offer/Answer model [5] for unicast and multicast usage, the limitations and rules described in [Section 8.2.2](#) of RFC YYYY [1] apply. Note that the profile\_idc byte of the H264-RCDO profile-level-id parameter can only take the value of 0 (no profile).

For interoperability with systems not supporting H264-RCDO, it is RECOMMENDED to offer the H264 media subtype as well. As specified in [RFC 3264](#) [5], listing the payload number for H264-RCDO before H264 in the format list on the "m=" line signals that H264-RCDO is preferred over H264. An example where this scheme is applied:

```
m=video 5555 RTP/AVP 97 98
a=rtpmap:97 H264-RCDO/90000
a=fmpdp:97 profile-level-id=008016;max-mbps=42000;max-smbps=323500
a=rtpmap:98 H264/90000
a=fmpdp:98 profile-level-id=428016;max-mbps=35000;max-smbps=323500
```

### 7.2. Declarative SDP Considerations

When H264-RCDO over RTP is offered with SDP in a declarative style, as in the Real Time Streaming Protocol (RTSP) [11] or the Session Announcement Protocol (SAP) [12], the considerations in [Section 8.2.3](#) of RFC YYYY [1] apply. Note that the profile\_idc byte of the H264-RCDO profile-level-id parameter can only take the value of 0 (no profile).



## **8. IANA Considerations**

This document requests that IANA registers H264-RCDO as specified in Section [Section 6.1](#). The media subtype is also requested to be added to the IANA registry for "RTP Payload Format MIME types" (<http://www.iana.org/assignments/rtp-parameters>).

## **9. Security Considerations**

RTP packets using the payload format defined in this specification are subject to the security considerations discussed in the RTP specification [\[6\]](#), and in any applicable RTP profile. Refer also to the security considerations of the RTP Payload Format for H.264 Video specification in RFC YYYY [\[1\]](#). No additional security considerations are introduced by this specification.

## **10. Acknowledgements**

The authors would like to acknowledge Gisle Bjoentegaard and Arild Fuldseth for their technical contribution to the specification. In the final phases Roni Even did a helpful review.

## **11. References**

### **11.1. Normative References**

- [1] Wang, Y., Even, R., Kristensen, T., and R. Jesup, "RTP Payload Format for H.264 Video", [draft-ietf-avt-rtp-rfc3984bis-12](#) (work in progress), October 2010.
- [2] International Telecommunications Union, "Advanced video coding for generic audiovisual services", ITU-T Recommendation H.264, November 2007.
- [3] International Telecommunications Union, "Extended video procedures and control signals for H.300-series terminals", ITU-T Recommendation H.241, May 2006.
- [4] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [5] Rosenberg, J. and H. Schulzrinne, "An Offer/Answer Model with Session Description Protocol (SDP)", [RFC 3264](#), June 2002.
- [6] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson,





- "RTP: A Transport Protocol for Real-Time Applications", STD 64, [RFC 3550](#), July 2003.
- [7] Schulzrinne, H. and S. Casner, "RTP Profile for Audio and Video Conferences with Minimal Control", STD 65, [RFC 3551](#), July 2003.
- [8] Handley, M., Jacobson, V., and C. Perkins, "SDP: Session Description Protocol", [RFC 4566](#), July 2006.
- [9] Josefsson, S., "The Base16, Base32, and Base64 Data Encodings", [RFC 4648](#), October 2006.
- [10] Casner, S., "Media Type Registration of RTP Payload Formats", [RFC 4855](#), February 2007.

### **11.2. Informative references**

- [11] Schulzrinne, H., Rao, A., and R. Lanphier, "Real Time Streaming Protocol (RTSP)", [RFC 2326](#), April 1998.
- [12] Handley, M., Perkins, C., and E. Whelan, "Session Announcement Protocol", [RFC 2974](#), October 2000.
- [13] Freed, N. and J. Klensin, "Media Type Specifications and Registration Procedures", [BCP 13](#), [RFC 4288](#), December 2005.
- [14] Lennox, J., Ott, J., and T. Schierl, "Source-Specific Media Attributes in the Session Description Protocol (SDP)", [RFC 5576](#), June 2009.

### **Authors' Addresses**

Tom Kristensen

TANDBERG

Philip Pedersens vei 22

N-1366 Lysaker

Norway

Phone: +47 67125125

Email: [tom.kristensen@tandberg.com](mailto:tom.kristensen@tandberg.com), [tomkri@ifi.uio.no](mailto:tomkri@ifi.uio.no)

URI: <http://www.tandberg.com>



Patrick Luthi  
TANDBERG  
Philip Pedersens vei 22  
N-1366 Lysaker  
Norway

Email: [patrick.luthi@tandberg.com](mailto:patrick.luthi@tandberg.com)

URI: <http://www.tandberg.com>